SCPISM: review, implementation and technical details

This section is a description of the SCPISM as implemented in the CHARMM program. This summary should be sufficient to implement the SCPISM into any molecular mechanics program.

In the SCPISM the electrostatic energy of a molecule composed of *N* atoms is given by

$$E_{elec} = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{D(r_{ij})r_{ij}} + \frac{1}{2}\sum_{i=1}^{N}\frac{q_i^2}{R_i}\left[\frac{1}{D(R_i)} - 1\right] \qquad (1)$$

where

$$D(r_{ij}) = \frac{1+\varepsilon_w}{1+k\exp(-\alpha_{ij}r_{ij})} - 1$$

$$D(R_i) = \frac{1+\varepsilon_w}{1+k\exp(-\alpha_i R_i)} - 1$$

are the screening functions, with $k = (\varepsilon_s - 1)/2$, ε_s is the static dielectric constant of the bulk solvent, and $\alpha_{ij} = (\alpha_i \alpha_j)^{1/2}$, where α_i is the screening parameter. These parameters depend on each atom *i* in the molecule, although the current implementation introduces only one parameter per chemical atom type as defined in the CHARMM topology files (*top_*.inp*).

In the SCPISM the effective Born radius R_i of an atom *i* in the macromolecule is given by $R_i = R_{i,w}\xi_{i,w} + R_{i,p}\xi_{i,p} + R_{i,A}\xi_{i,A}$, with $\xi_{i,w} + \xi_{i,p} + \xi_{i,A} = 1$, where $\xi_{i,x}$ is the fraction of atom *i* exposed to component *x* (*w*=solvent; *p*=solute; *A*=acceptor) so $R_{i,x}$ can be thought of as the Born radius of atom *i* if fully imbedded in *x*. For atoms other than polar hydrogens $\xi_{i,A} = 0$ by definition and, then, $R_i = R_{i,w}\xi_{i,w} + R_{i,p}\xi_{i,p}$, with $\xi_{i,w} + \xi_{i,p} = 1$. With this definition the Born radius changes with the conformation of the system.

$R_{i,w}$ is defined by $R_{i,w} = R_{i,COV} + \delta_i(q_i)$, where $R_{i,COV}$ is the covalent radius of atom *i* and $\delta_i(q_i)$ is an extension that depends on the atom type in the molecule and on its partial

charge q_i. In the current implementation the quantity $\delta_i(q_i)$ depends only on the sign of the charge, not on its magnitude, and is independent on the atom, i.e., $\delta_i(q_i) \equiv \delta(\text{sign}(q_i))$, and given by $\delta(+) = 0.35$ Å and $\delta(-) = 0.85$ Å (this is similar to Latimer's approach of ion solvation); five values of $R_{i,COV}$ are introduced, one for each chemical element C, O, N, S, and H.

In analogy with $R_{i,w}$, the radius $R_{i,p}$ is defined as $R_{i,p} = R_{i,COV} + \gamma_i(q_i)$, where γ is the extension of the atom in its particular molecular environment. In the current implementation γ is approximated as $\gamma_i(q_i) = \delta(sign(q_i)) + \lambda$, where $\lambda = 0.5$ Å for all atoms in the molecule.

The values of $R_{i,A}$ are as important for the quality of the results as the screening parameters α_i that controls bulk electrostatics (defined in $D(x)$ above). These coefficients allow for fine-tuning the hydrogen bonding interactions in the system. Unlike $R_{i,w}$ and $R_{i,p}$ that are calculated based on the definitions above, $R_{i,A}$ are empirically adjusted to target the hydrogen-bonding (HB) energy between the donor and the acceptor group that share the proton *i*. This adjustment is necessary because the screening functions $D(r)$ are based on bulk properties and contain no information on the rearrangement of the solvent structure around a donor and an acceptor when these groups are close to each other. The presence of a donor groups in the vicinity of an acceptor not only affects the accessibility of solvent around the acceptor (and donor) by removing its hydration shells (what is accounted for by the Born radii through $\xi_{i,x}$), but it also modifies the structure of the hydration shell in particular ways. This specific rearrangement of the solvent structure creates so-called solvent induced forces (SIF), e.g., in the form of water molecules bridging the acceptor and donor groups, that affects the interaction forces between these groups, hence the effective HB interaction energies. Controlling the strength of the HB interactions by adjusting the Born radii of the shared proton is a practical and very efficient way of accounting for these competing effects (this is similar to how CHARMM

fine tunes HB strength in vacuum, i.e., by adjusting, in particular, the vdW radii of the shared proton).

Although the fractions $\xi_{i,x}$ can be calculated numerically the version of the SCPISM for MD simulations uses a contact-like model approximation, i.e.,

$$\xi_{i,w} = \left[A_i - \sum_{j \neq i}^{N} B_{ij} \exp(-C_{ij} r_{ij}) \right] \Big/ 4\pi (R_p + R_{i,vdW})^2 \qquad (2)$$

where R_p and $R_{i,vdW}$ are the probe radius (equal to 1.4Å) and the van der Waals radius of atom *i*, respectively. Calculating $\xi_{i,x}$ based on a contact model is desirable since it gives partial information on how deep atom *i* is within the protein (this information would be missing if these fraction were based on exact calculation of solvent accessible surface areas). Here a simplification is made where the coefficients B_{ij} and C_{ij} depend only on the central atom *i*, i.e., $B_{ij} \equiv B_i$ and $C_{ij} \equiv C_i$. For non-polar hydrogen atoms the fraction $\xi_{i,p}$ is obtained from $\xi_{i,p} = 1 - \xi_{i,w}$. For a polar hydrogen the fraction exposed to the acceptor atoms, $\xi_{i,A}$, is subtracted from the fraction exposed to the solute, $\xi_{i,p}$, i.e.,

$$\xi_{i,p} = 1 - \xi_{i,w} - \sum_{j \neq i}^{M} D_{ij} \exp(-E_{ij} r_{ij}) \qquad (3)$$

$$\xi_{i,A} = \sum_{j \neq i}^{M} D_{ij} \exp(-E_{ij} r_{ij}) \qquad (4)$$

and $\xi_{i,w}$ is still defined by Eq.(2). In Eqs.(3) and (4) the sums run over all the acceptor atoms (*M* is the total number of acceptors); the coefficients depend only on the central atom *i*, i.e., $D_{ij} \equiv D_i$ and $E_{ij} \equiv E_i$. In practice, the summations in Eqs.(2)-(4) are restricted to atoms within a cutoff distance r_{cutoff} of the central atom *i*, which in the current implementation is set to $r_{cutoff} = 6$ Å.

Coefficients A_i, B_i, C_i, D_i and E_i in Eqs.(2)-(4) are optimized as described in previous publications. These coefficients are not part of the SCPISM *per se* but a complement to it, that account for a particular aspect of the topology of the molecule.

Values of parameters α_i (electrostatics) and $R_{i,A}$ (hydrogen bonding) depend on the force field used (for the current implementation in CHARMM version c31b1 see *scpism.inp* parameter file).

The polar component of the solvation energy in the SCPISM is given by

$$\Delta G_{elec} = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{r_{ij}}\left[\frac{1}{D(r_{ij})}-\frac{1}{D_0(r_{ij})}\right]+ \\ +\frac{1}{2}\sum_{i=1}^{N}q_i^2\left\{\frac{1}{R_i}\left[\frac{1}{D(R_i)}-1\right]-\frac{1}{R_i}\left[\frac{1}{D_0(R_i)}-1\right]\right\} \tag{5}$$

where $D_0(x)$ is the screening function defined in the system when the molecule is in the vacuum in the same conformation. In the SCPISM the screenings $D_0(x)$ are assumed to be constant, i.e., $D_0(x) = D_I$ where the index I indicates the side chain analogs used in the parameters optimization (see below). In this case, each D_I should be equal to the static dielectric constant ε_I of the corresponding analog in the gas phase, which is dominated mainly by the electron structure. Values for ε_I can be obtained from *ab initio* methods or from available experimental data. However, for the parameter optimization within an non-polarizable, point-charge-based, empirical force field like CHARMM it seems reasonable to treat $\{D_I\}$ as open parameters on the same foot as $\{\alpha_i\}$.

Parameters Optimization:

With the approximation $D_0(x) = D_I$, the polar contribution to the free energy of each analog I, is expressed as

$$\Delta G^{I}_{elec} = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j}{r_{ij}}\left[\frac{1}{D(r_{ij})}-\frac{1}{D_I}\right]+ \\ +\frac{1}{2}\sum_{i=1}^{N}q_i^2\left\{\frac{1}{R_i}\left[\frac{1}{D(R_i)}-1\right]-\frac{1}{R_{vdW,I}}\left[\frac{1}{D_I}-1\right]\right\} \tag{6}$$

The parameter optimization in the SCPISM was based on a simulated annealing Monte Carlo in the space of the parameters $\{\alpha_i, D_I\}$. A function Γ is defined by

$$\Gamma(\{\alpha_i\};\{D_I\}) = \sum_I \left[\Delta G^I_{elec}(\{\alpha_i\};\{D_I\}) - \Delta G^I_{elec,0}\right]^2 \quad (7)$$

where $\Delta G^I_{elec,0}$ is the polar component of the experimental hydration energy of each molecule. Thus, a simulated annealing MC technique can be used with a standard Metropolis criterion in the canonical ensemble if a Boltzmann factor $f = exp(-\Gamma/T)$ is defined; T is an arbitrary parameter that decreases according to a standard logarithmic schedule. Stochastic sampling in the space $\{\alpha_i, D_I\}$ leads to a minimization of Γ (ideally $\Gamma \rightarrow \Gamma_0 = 0$; in practice $\Gamma_0 > 0$).